BU-933-M

CORNELL
UNIVERSITY
MATHEMATICAL
SCIENCES
INSTITUTE

ANNOTATED COMPUTER OUTPUT

'87 - 8

ANNOTATED COMPUTER OUTPUT FOR SPLIT PLOT DESIGN: SAS GLM

by

W.T. FEDERER, Z.D. FENG, M.P. MEREDITH, AND N.J. MILES-MCDERMOTT

November 1987

201 Caldwell Hall ■ Ithaca, New York 14853-2602

(607) 255-8005

ANNOTATED COMPUTER OUPUT FOR SPLIT PLOT DESIGN: SAS GLM

by

W.T. Federer, Z.D. Feng, M.P. Meredith, and N.J. Miles-McDermott

Mathematical Sciences Institute, Cornell University, Ithaca, NY

ABSTRACT

The analysis of covariance for split plot designs is not always straightforward when using a statistical software package such as SAS PROC GLM. In order to demonstrate correct analyses several data sets are examined and annotated SAS output is given. Hypothetical data are analyzed first without and then with the covariate included. The whole plots are arranged in a RCBD and the covariate is measured on the subplot experimental units. A second example has whole plots arranged in a CRD and the covariate measured only on the whole plot experimental units.

Complete ANOVA tables for both examples may be computed in a single procedural call to SAS PROC GLM. Both Type I and Type III sums of squares are necessary to construct the proper ANOVA table. A commonly employed approach requiring two separate procedural calls to GLM is also demonstrated. Formulae for the standard errors of the difference between adjusted whole plot and subplot means are reported.

INTRODUCTION

This is part of a continuing project that produces annotated computer output for the analysis of balanced split plot experiments with covariates. The complete project will involve processing three examples on SAS/GLM, BMDP/2V, SPSS-X/MANOVA, GENSTAT/ANOVA, and SYSTAT/MGLH. Only univariate results are considered. We show here the results from SAS GLM.

For Example 1, the data are artificial and were constructed for ease of computation; the experiment design for the whole plots is a randomized complete block and the split plot treatments are randomly allocated to the split plot experimental units within <u>each</u> whole plot. Example 2 is the same as Example 1 except that a covariate varies from split plot to split plot. The data for Example 3 come from an experiment wherein the whole plot treatments are laid out in a completely randomized design and the split plot treatments are randomly allotted to the split plot experimental units within <u>each</u> whole plot. The value of the covariate varies from whole plot to whole plot but is constant for all split plots within a whole plot treatment.

We present the elementary computational steps. Simple hypothetical data are used for the first two examples so that it is easy to provide all detailed computations to illustrate how each number is obtained. Some readers may wish to skip the detailed computations (see Federer, 1955, Chapter XVI). The third example comes from Winer (1971). The detailed computations are given in his book (p. 803).

Data SP-1

Split plot data with whole plots arranged in randomized complete block design (hypothetical data)

	Whole plot treatment									
	W1					W2				
Block	split plot treatment S_1	S_2	S_3	S_4	Total	split plot treatment S_1	S_2	S_3	S_4	Total
1	3	4	7	6	20	3	2	1	14	20
2	6	10	1	11	28	8	8	2	18	36
3	6	10	4	4	24	10	8	9	13	40
Total	15	24	12	21	72	21	18	12	45	96

Total and Means

Blocks (8 observations)	Total	Mean	W(whole plots) (12 observations)	Total	Mean	S(split plot) (6 observations)	Total	Mean
1	40	5	W1	72	6	S_1	36	6
2	64	8	W2	96	8	S_2	42	7
3	64	8				S_3	24	4
Grand Total		168				S_4	66	11
Grand Mean		7						

Model: $Y_{ijk} = \mu + \rho_j + \tau_i + \delta_{ij} + \alpha_k + (\alpha\tau)_{ik} + \epsilon_{ijk}$

μ = mean

ρ_j = effect of block j

δ_{ij} = error (a)

ϵ_{ijk} = error (b)

τ_i = effect of whole plot i

α_k = effect of split plot k

$(\alpha\tau)_{ik}$ = effect of interaction of whole plot i and split plot k

where it is assumed that $\rho_j \sim N(o,\sigma_\rho^2)$, $\delta_{ij} \sim N(o,\sigma_\delta^2)$, $\epsilon_{ijk} \sim N(o,\sigma_\epsilon^2)$, and ρ_j, δ_{ij}, and ϵ_{ijk} are mutually independent. $i=1,2,\cdots,a$, $j=1,2,\cdots,r$, and $k=1,2,\cdots,s$.

Analysis of Variance

Source	(*)	df	SS
B (Blocks)	$= R(\rho \mid \mu,\tau,\alpha,\alpha\tau)$	2	48
W (whole plot treatments)	$= R(\tau \mid \mu,\rho,\alpha,\alpha\tau)$	1	24
B×W (error (a))	$= R(\delta \mid \mu,\rho,\tau,\alpha,\alpha\tau)$	2	16
S (split plot treatments)	$= R(\alpha \mid \mu,\rho,\tau,\alpha\tau)$	3	156
S×W (interaction of S and W)	$= R(\alpha\tau \mid \mu,\alpha,\tau,\rho)$	3	84
(**) S×B:W (error (b))	$= R(\epsilon \mid \mu,\alpha,\tau,\alpha\tau,\rho)$	12	112
Total (Corrected for mean)	$= R(\rho,\tau,\delta,\alpha,\alpha\tau,\epsilon \mid \mu)$	23	440
Mean	$= R(\mu)$	1	1176
Total (Uncorrected for mean)	$= R(\mu,\rho,\tau,\delta,\alpha,\alpha\tau,\epsilon)$	24	1616

(*)Notation follows that of Searle(1971); since the design is balanced, $R(\rho \mid \mu,\tau,\alpha,\alpha\tau) = R(\rho \mid \mu)$, etc. The simpler notation is used later.
(**) S×B:W means S×B within W.

Calculations of SS's:

$N = 2\cdot 3\cdot 4 = 24 \quad , \quad \bar{Y} = 7$

$$R(\mu,\rho,\tau,\delta,\alpha,\alpha\tau,\epsilon) = \sum_{i=1}^{2}\sum_{j=1}^{3}\sum_{k=1}^{4} Y_{ijk}^2 = (3^2+6^2+6^2+\cdots+18^2+13^2) = 1616$$

$$R(\mu) = N\bar{Y}^2 = 24\cdot(7)^2 = 1176$$

$$R(\rho,\tau,\delta,\alpha,\alpha\tau,\epsilon \mid \mu) = 1616 - 1176 = 440$$

$$R(\rho \mid \mu) = R(\mu,\rho) - R(\mu) = \frac{(40^2+64^2+64^2)}{8} - 1176 = 1224 - 1176 = 48$$

$$R(\tau \mid \mu) = R(\mu,\tau) - R(\mu) = \frac{(72^2+96^2)}{12} - 1176 = 1200 - 1176 = 24$$

$$R(\delta \mid \mu,\rho,\tau) = R(\delta,\mu,\rho,\tau) - R(\mu,\rho) - R(\tau,\mu) + R(\mu)$$

$$= \frac{(20^2+28^2+24^2+20^2+36^2+40^2)}{4} - 1224 - 1200 + 1176$$

$$= 1264 - 1224 - 1200 + 1176 = 16$$

$$R(\alpha \mid \mu) = R(\alpha,\mu) - R(\mu) = \frac{(36^2+42^2+24^2+66^2)}{6} - 1176 = 1332 - 1176 = 156$$

$$R(\alpha\tau \mid \mu,\alpha,\tau) = R(\alpha\tau,\mu,\alpha,\tau) - R(\mu,\alpha) - R(\mu,\tau) + R(\mu)$$

$$= \frac{(15^2+24^2+12^2+21^2+21^2+18^2+12^2+45^2)}{3} - 1332 - 1200 + 1176$$

$$= 1440 - 1332 - 1200 + 1176 = 84$$

$$R(\epsilon|\mu,\rho,\delta,\alpha,\tau,\alpha\tau) = R(\epsilon,\mu,\alpha,\rho,\delta,\tau,\alpha\tau) - R(\mu,\rho,\tau,\delta) - R(\mu,\alpha,\tau,\alpha\tau) + R(\tau,\mu)$$

$$= 1616 - 1264 - 1440 + 1200 = 112$$

Data SP-2

Data SP-2: Data SP-1 with the following covariate Z which varies with split plot

Covariate (Z)

	whole plot									
	W1					W2				
	S_1	S_2	S_3	S_4	Total	S_1	S_2	S_3	S_4	Total
B_1	1	2	1	2	6	2	0	2	4	8
B_2	2	2	0	4	8	4	1	3	4	12
B_3	3	5	2	0	10	3	2	4	7	16
Total	6	9	3	6	24	9	3	9	15	36

Totals and Means

blocks (8 observations)	Total	Mean	W (whole plot) (12 observations)	Total	Mean	S (split plot) (6 observations)	Total	Mean
1	14	14/8	1	24	2.0	1	15	2.5
2	20	20/8	2	36	3.0	2	12	2.0
3	26	26/8				3	12	2.0
Grand						4	21	3.5
Total	60	2.5						

Model: $Y_{ijk} = \mu+\rho_j+\tau_i+\delta_{ij}+\alpha_k+(\alpha\tau)_{ik}+\beta_1(\bar{Z}_{ij.}-\bar{Z}_{...})+\beta_2(Z_{ijk}-\bar{Z}_{ij.})+\epsilon_{ijk}$

β_1 = whole plot regression slope β_2 = split plot regression slope

where μ, ρ_j, τ_i, δ_{ij}, α_k, $(\alpha\tau)_{ik}$, and ϵ_{ijk} are as in SP-1, and $\bar{Z}_{ij.}$ and $\bar{Z}_{...}$ are the arithmetic means for Z_{ijk} .

Table of sum of squares and cross products

Source	df	YY	YZ	ZZ
B	2	48	18	9
W	1	24	12	6
B×W (error a)	2	16	4	1
S	3	156	33	9
S×W	3	84	33	21
S×B:W (error b)	12	112	17	20
Mean	1	1176	420	150
Total	24	1616	537	216

YY column is the same as in SP-1, ZZ column is computed in the same fashion. Thus, only computations for YZ column are illustrated.

$$\text{Total}_{YZ} = \sum_{i=1}^{2} \sum_{j=1}^{3} \sum_{k=1}^{4} Y_{ijk} \cdot Z_{ijk}$$

$$= 3(1) + 6(2) + \cdots + 14(4) + 18(4) + 13(7) = 537$$

$$\text{Mean}_{YZ} = N\bar{Y}_{\cdots}\bar{Z}_{\cdots} = \frac{168 \cdot 60}{24} = 420$$

$$B_{YZ} = \frac{\sum_{j=1}^{3} \left(\sum_{i=1}^{2} \sum_{k=1}^{4} Y_{ijk}\right)\left(\sum_{i=1}^{2} \sum_{k=1}^{4} Z_{ijk}\right)}{2 \cdot 4} - 420 = \frac{40(14)+64(20)+64(26)}{8} - 420$$

$$= 438 - 420 = 18$$

$$W_{YZ} = \frac{\sum_{i=1}^{2} \left(\sum_{j=1}^{3} \sum_{k=1}^{4} Y_{ijk}\right)\left(\sum_{j=1}^{3} \sum_{k=1}^{4} Z_{ijk}\right)}{3(4)} - 420 = 432 - 420 = 12$$

$$B{\times}W_{YZ} = \frac{\sum_{i=1}^{2} \sum_{j=1}^{3} \left(\sum_{k=1}^{4} Y_{ijk}\right)\left(\sum_{k=1}^{4} Z_{ijk}\right)}{4} - 438 - 432 + 420$$

$$= 454 - 438 - 432 + 420 = 4$$

$$S_{YZ} = \sum_{k=1}^{4} \frac{\left(\sum_{i=1}^{2} \sum_{j=1}^{3} Y_{ijk}\right)\left(\sum_{i=1}^{2} \sum_{j=1}^{3} Z_{ijk}\right)}{2(3)} - 420 = 453 - 420 = 33$$

$$S{\times}W_{YZ} = \frac{\sum_{i=1}^{2} \sum_{k=1}^{4} \left(\sum_{j=1}^{3} Y_{ijk}\right)\left(\sum_{j=1}^{3} Z_{ijk}\right)}{3} - 453 - 432 + 420$$

$$= 498 - 453 - 432 + 420 = 33$$

$$S\times B{:}W_{YZ} = \sum_{i=1}^{2}\sum_{j=1}^{3}\sum_{k=1}^{4} Y_{ijk}Z_{ijk} - 454 - 498 + 432$$

$$= 537 - 454 - 498 + 432 = 17$$

Analysis of Variance and Covariance

Source		df	SS
B (block)	$= R(\rho\|\mu,\tau)$	2	48
W (whole plot treatment)	$= R(\tau\|\mu,\rho,\beta_1)$	1	3.4286
Regression (a)	$= R(\beta_1\|\mu,\rho,\tau)$	1	16.0
B×W (error (a))	$= R(\delta\|\mu,\rho,\tau,\beta_1)$	1	0.0
S (split plot treatment)	$= R(\alpha\|\mu,\rho,\tau,\alpha\tau,\beta_2)$	3	84.243
S×W (interaction of S and W)	$= R(\alpha\tau\|\mu,\rho,\tau,\alpha,\beta_2)$	3	37.474
Regression (b)	$= R(\beta_2\|\mu,\rho,\tau,\alpha,\alpha\tau)$	1	14.450
S×B: W (error (b))	$= R(\epsilon\|\mu,\rho,\alpha,\tau,\alpha\tau,\beta_2)$	11	97.550
Total (corrected for mean)		23	440

$$\hat{\beta}_1 = B\times W_{YZ} \;/\; B\times W_{ZZ} = 4/1 = 4$$

$$\hat{\beta}_2 = S\times B{:}W_{YZ} \;/\; S\times B{:}W_{ZZ} = 17/20 = 0.85$$

The SS's adjusted by regression on Z are illustrated below:

$R(\rho|\mu)$ = 48, remains same since it is not of interest to adjust blocks for Z.

$$R(\tau,\delta|\mu,\rho,\beta_1) = (W_{YY} + B\times W_{YY}) - \frac{(W_{YZ} + B\times W_{YZ})^2}{W_{ZZ} + B\times W_{ZZ}}$$

$$= (24 + 16) - \frac{(12 + 4)^2}{6 + 1} = 40 - \frac{256}{7} = 3.4286$$

$$R(\delta|\mu,\rho,\tau,\beta_1) = B\times W_{YY} - \frac{(B\times W_{YZ})^2}{B\times W_{ZZ}} = 16 - \frac{4^2}{1} = 0$$

$$R(\tau|\mu,\rho,\beta_1) = R(\tau,\delta|\mu,\rho,\beta_1) - R(\delta|\mu,\rho,\tau,\beta_1)$$

$$= 40 - \frac{256}{7} - 0 = 3.4286$$

$$R(\beta_1|\mu,\tau,\rho) = \frac{(B{\times}W_{YZ})^2}{B{\times}W_{ZZ}} = \frac{4^2}{1} = 16$$

$$R(\alpha,\epsilon|\mu,\rho,\tau,\alpha\tau,\beta_2) = (S_{YY} + S{\times}B{:}W_{YY}) - \frac{(S_{YZ} + S{\times}B{:}W_{YZ})^2}{S_{ZZ} + S{\times}B{:}W_{ZZ}}$$

$$= (156 + 112) - \frac{(33+17)^2}{9+20}$$

$$= 268 - 86.207 = 181.793$$

$$R(\alpha\tau,\epsilon|\mu,\rho,\alpha,\tau,\beta_2) = (S{\times}W_{YY} + S{\times}B{:}W_{YY}) - \frac{(S{\times}W_{YZ} + S{\times}B{:}W_{YZ})^2}{S{\times}W_{ZZ} + S{\times}B{:}W_{ZZ}}$$

$$= 84 + 112 - \frac{(33+17)^2}{21+20} = 196 - 60.976 = 135.024$$

Note: $R(\alpha,\epsilon|\mu,\beta_2)$ and $R(\alpha\tau,\epsilon|\mu,\alpha,\tau,\beta_2)$ are intermediate steps for later use.

$$R(\beta_2|\mu,\rho,\alpha,\tau,\alpha\tau) = \frac{(S{\times}B{:}W_{YZ})^2}{S{\times}B{:}W_{ZZ}} = \frac{17^2}{20} = 14.450$$

$$R(\epsilon|\mu,\rho,\alpha,\tau,\alpha\tau,\beta_2) = S{\times}B{:}W_{YY} - \frac{(S{\times}B{:}W_{YZ})^2}{S{\times}B{:}W_{ZZ}} = 112 - \frac{17^2}{20} = 112 - 14.45 = 97.55$$

$$R(\alpha|\mu,\rho,\tau,\alpha\tau,\beta_2) = R(\alpha,\epsilon|\mu,\rho,\tau,\alpha\tau,\beta_2) - \text{SS error b} = 181.793 - 97.55$$
$$= 84.243$$

$$R(\alpha\tau|\mu,\rho,\alpha,\tau,\beta_2) = R(\alpha\tau,\epsilon|\mu,\rho,\alpha,\tau,\beta_2) - R(\epsilon|\mu,\rho,\alpha,\tau,\alpha\tau,\beta_2)$$

$$= 135.024 - 97.55 = 37.474$$

Data SP-3

Split plot data with plots arranged in a completely randomized design and a covariate Z that is constant within the whole plot. (Winer, 1971, p. 803)

whole plot	Subject	Split plots B_1 Y	B_2 Y	Z	Total Y
A_1					
	1	10	8	3	18
	2	15	12	5	27
	3	20	14	8	34
	4	12	6	2	18
A_2					
	5	15	10	1	25
	6	25	20	8	45
	7	20	15	10	35
	8	15	10	2	25
	Total	132	95	39	227
	Mean	16.5	11.9	4.88	

Model: $y_{ijk} = \mu + \tau_i + \delta_{ij} + \alpha_k + (\tau\alpha)_{ik} + \beta_1(Z_{ij}-\bar{Z}_{..}) + \epsilon_{ijk}$

τ_i = A effect (whole plot) $\qquad \delta_{ij}$ = error (a) $\qquad \epsilon_{ijk}$ = error (b)

α_k = B effect (split plot) $\qquad \beta_1$ = whole plot regression slope

where $\delta_{ijk} \sim N(o,\sigma_\delta^2)$, $\epsilon_{ijk} \sim N(o,\sigma_\epsilon^2)$, δ_{ij} and ϵ_{ijk} are mutually independent. $i=1,2,\cdots,a$, $j=1,2,\cdots,r$, and $k=1,2,\cdots,s$.

Analysis of variance and covariance

Source		df	SS
A (whole plot)	$= R(\tau\|\mu,\beta_1)$	1	44.492
Regression	$= R(\beta_1\|\mu,\tau)$	1	166.577
Error (a)	$= R(\delta\|\mu,\tau,\beta_1)$	5	61.298
B (split plot)	$= R(\alpha\|\mu,\tau,\alpha\tau)$	1	85.563
A×B (interaction)	$= R(\tau\alpha\|\mu,\tau,\alpha)$	1	0.563
Error (b)	$= R(\epsilon\|\mu,\tau,\alpha,\tau\alpha)$	6	6.375
Total (corrected)	$= R(\tau,\alpha,\beta_1,\tau\alpha,\delta\|\mu)$	15	388.438

Table of SS and products

Symbol	Y^2	ZY	Z^2
W	68.06	12.38	2.75
E(a)	227.88	163.00	159.50
S	85.563	0	0
WS	0.563	0	0
E(b)	6.375	0	0

$$\hat{\beta}_1 = \frac{163.00}{159.50} = 1.02$$

Since the computations are illustrated in Winer (1971, p. 803-5) we have omitted them here.

Many SAS users would likely adopt an analysis of covariance strategy for split plot designs that requires two procedural calls - one for the whole plot analysis and another for the split plot analysis. These analyses are presented under SP-2 and SP-3. However, it is possible to obtain the complete ANOVA tables for SP-2 and SP-3 in a single procedural call of SAS GLM. This latter approach is recommended and is given in SP-2A and SP-3A.

SP-1: Control Language

Control language is typed in upper case and comments are bolded.

```
DATA ONE;
INPUT BLOCK WHOLE SUBPLOT Y;     ⇒ Input variables
TITLE SP-1:  SPLIT PLOTS WITH WHOLE PLOTS ARRANGED IN RCB DESIGN;
CARDS;      ⇒ Tells SAS that data follow
1 1 1 3
1 1 2 4
1 1 3 7
:           ⇒ Data are entered with only one datum per line
:
:
3 2 4 13
PROC GLM;
CLASS BLOCK WHOLE SUBPLOT;  ⇒  Designates classification variables
MODEL YIELD=BLOCK WHOLE BLOCK*WHOLE
WHOLE WHOLE*SUBPLOT/SS3 P;  ⇒  Designates model being used.  The
                               SS3 option requests only type III sums
                               of squares and P requests residuals
                               (only one type SS's was requested
                               because the data are balanced making
                               all types SS's equal).  Type I SS's are
                               the cheapest to compute.
TEST H=BLOCK WHOLE E=BLOCK*WHOLE;  ⇒  Requests SAS to test the whole
                                      plot effects using error(a)
```

Note: SAS always computes F tests based on the residual sum of squares. This is not always the appropriate test in split plot analyses so adding the TEST statement (above) is critical to obtaining an appropriate test.

SP-2: Control Language

Note: Because estimates of both a whole plot regression slope and split plot regression slope are needed, two procedural calls to SAS GLM are required. The first call gives the appropriate whole plot analysis and the second gives the appropriate split plot analysis.

Procedural Call for Whole Plot Analysis

```
DATA ONE;
TITLE1 SP-2:  SPLIT PLOT DESIGN WITH WHOLE PLOTS ARRANGED IN RCB:
TITLE2        WITH A COVARIATE VARYING WITH SPLIT PLOT;
INPUT BLOCK WHOLE Z1 Z2 Z3 Z4 Y1 Y2 Y3 Y4;
Z=(SUM(OF Z1-Z4)/2);  ⇒  Z and Y are scaled for this analysis
Y=(SUM(OF Y1-Y4)/2);     so that the sums of squares are correct
CARDS;
1 1 1 2 1 2 3 4 7 6
2 1 2 2 0 4 6 10 1 11    ⇒  For whole plot analysis data must be
3 1 3 5 2 0 6 10 4 4         organized in a similar arrangement
1 2 2 0 2 4 3 2 1 14         with all split plot values for a
2 2 4 1 3 4 8 8 2 18         particular BLOCK by WHOLE combination
3 2 3 2 4 7 10 8 9 13        on the same line (see INPUT statement
PROC GLM;                    above for order)
TITLE3 CORRECT WHOLE PLOT ANALYSIS;
CLASS BLOCK WHOLE;
MODEL Y=BLOCK WHOLE Z/SOLUTION SS1 SS3 P; ⇒ The SOLUTION option yields
                                            the parameter estimates and
                                            so gives the estimated
                                            regression slope
LSMEANS BLOCK WHOLE/STDERR;      ⇒  Yields adjusted treatment means and
                                     standard errors.
ESTIMATE 'WHOLE PLOT SLOPE' Z1;  ⇒  The ESTIMATE statement gives the
                                     estimated regression slope and its
                                     standard error directly.
```

Procedural Call for Split Plot Analysis

```
DATA TWO;
INPUT BLOCK WHOLE SUBPLOT Z Y;
CARDS;
1 1 1 1 3
1 1 2 2 4
1 1 3 1 7
:
:
:
3 2 4 7 13
PROC GLM;
TITLE3 'CORRECT SPLIT PLOT ANALYSIS';
CLASS BLOCK WHOLE SUBPLOT;
MODEL Y=BLOCK WHOLE BLOCK*WHOLE SUBPLOT
WHOLE*SUBPLOT Z/SOLUTION SS1 SS3 P;
LSMEANS SUBPLOT WHOLE*SUBPLOT;
ESTIMATE 'SUBPLOT SLOPE' Z 1;
```

SP-3: Control Language

Note: Even though we are estimating only one slope in this example, two procedural calls are required in order to estimate the regression slope. If both the whole plot and split plot are specified in one run, Z becomes confounded in SUB(A) and β_1 cannot be estimated correctly.

Procedural Call for Correct Whole Plot Analysis

```
DATA ONE;
INPUT A Z Y1 Y2;
SUBJECT =_N_;
MY=(SUM(OF Y1-Y2))/(SQRT(2));   ⟹  Z and Y rescaled so SS's agree with
                                     those of Winer
Z = 2*Z/(SQRT(2));
TITLE1 SP-4:  SPLIT PLOT DESIGN WITH WHOLE PLOTS ARRANGED IN CRD;
TITLE2        WITH A COVARIATE CONSTANT IN SPLIT PLOT;
CARDS;
1 3 10 8
1 5 15 12
1 8 20 14
:
:
:
2 2 15 10
PROC GLM;
CLASS A;
MODEL MY=Z A/SOLUTION SS1 SS3 P;
TITLE3 CORRECT WHOLE PLOT ANALYSIS;
LSMEANS A/STDERR;
ESTIMATE 'REGR SLOPE' Z 1;
```

Procedural Call for Correct Split Plot Analysis

```
DATA TWO;
INPUT SUB A B Y;
CARDS;
1 1 1 10
1 1 2  8
2 1 1 15
:
:
:
8 2 2 10
PROC GLM;
CLASS SUB A B;
MODEL Y=A SUBJECT(A) B B*A/SS1 SS3 P;
TITLE3 CORRECT SPLIT PLOT ANALYSIS;
LSMEANS B B*A/STDERR;
```

Variances and Standard Errors of Adjusted Means and Differences Amongst Adjusted Means for SP-2

$$\mathrm{Var}(\bar{Y}_{i\cdot k\ adj}) = (\sigma^2_\rho + \sigma^2_\delta + \sigma^2_\epsilon)/r + (\sigma^2_\epsilon + s\sigma^2_\delta)(\bar{Z}_{i\cdot\cdot} - \bar{Z}_{\cdot\cdot\cdot})^2/W{\times}B_{ZZ}$$

$$+ \sigma^2_\epsilon\ (\bar{Z}_{i\cdot k} - \bar{Z}_{i\cdot\cdot})^2/S{\times}B{:}W_{ZZ}$$

$$\mathrm{Var}(\bar{Y}_{i\cdot\cdot\ adj}) = [\sigma^2_\epsilon + s(\sigma^2_\rho + \sigma^2_\delta)]/rs + (\sigma^2_\epsilon + s\sigma^2_\delta)(\bar{Z}_{i\cdot\cdot} - \bar{Z}_{\cdot\cdot\cdot})^2/W{\times}B_{ZZ}$$

$$\mathrm{Var}(\bar{Y}_{\cdot\cdot k\ adj}) = (\sigma^2_\rho + \sigma^2_\delta + \sigma^2_\epsilon)/ar + \sigma^2_\epsilon\ (\bar{Z}_{\cdot\cdot k} - \bar{Z}_{\cdot\cdot\cdot})^2/S{\times}B{:}W_{ZZ}$$

$$\mathrm{Var}(\bar{Y}_{i\cdot\cdot\ adj} - \bar{Y}_{i'\cdot\cdot\ adj}) = (\sigma^2_\epsilon + s\sigma^2_\delta)\left[\frac{2}{rs} + \frac{(\bar{Z}_{i'\cdot\cdot} - \bar{Z}_{i\cdot\cdot})^2}{W{\times}B_{ZZ}}\right]$$

$$\mathrm{Var}(\bar{Y}_{\cdot\cdot k\ adj} - \bar{Y}_{\cdot\cdot k'\ adj}) = \sigma^2_\epsilon\left[\frac{2}{ar} + \frac{(\bar{Z}_{\cdot\cdot k'} - \bar{Z}_{\cdot\cdot k})^2}{S{\times}B{:}W_{ZZ}}\right]$$

$$\mathrm{Var}(\bar{Y}_{i\cdot k\ adj} - \bar{Y}_{i\cdot k'\ adj}) = \sigma^2_\epsilon\left[\frac{2}{r} + \frac{(\bar{Z}_{i\cdot k'} - \bar{Z}_{i\cdot k})^2}{S{\times}B{:}W_{ZZ}}\right]$$

and, for $i \neq i'$

$$\mathrm{Var}(\bar{Y}_{i\cdot k\ adj} - \bar{Y}_{i'\cdot k'\ adj}) = \frac{2}{r}\ (\sigma^2_\epsilon + \sigma^2_\delta) + \frac{(\bar{Z}_{i'\cdot\cdot} - \bar{Z}_{i\cdot\cdot})^2}{W{\times}B_{ZZ}}\ (\sigma^2_\epsilon + s\sigma^2_\delta)$$

$$+ \frac{(\bar{Z}_{i'\cdot k'} - \bar{Z}_{i'\cdot\cdot} - \bar{Z}_{i\cdot k} + \bar{Z}_{i\cdot\cdot})^2}{S{\times}B{:}W_{ZZ}}\ \sigma^2_\epsilon$$

Estimates of the variance components σ^2_ϵ and σ^2_δ are required to calculate standard errors of the above differences amongst adjusted treatment means. From the expected mean squares of the ANOVA table it is known that error(a) and error(b) estimate $\sigma^2_\epsilon + s\sigma^2_\delta$ and σ^2_ϵ, respectively. If error(a) and error(b) are denoted E_a and E_b, respectively, then σ^2_δ is estimated by $(E_a - E_b)/s$. Hence, the desired standard errors are given by:

$$SE(\bar{Y}_{i\cdot\cdot\ adj} - \bar{Y}_{i'\cdot\cdot\ adj}) = \sqrt{E_a\left[\frac{2}{rs} + \frac{(\bar{Z}_{i'\cdot\cdot} - \bar{Z}_{i\cdot\cdot\cdot})^2}{W\times B_{ZZ}}\right]}$$

$$SE(\bar{Y}_{\cdot\cdot k\ adj} - \bar{Y}_{\cdot\cdot k'\ adj}) = \sqrt{E_b\left[\frac{2}{ar} + \frac{(\bar{Z}_{\cdot\cdot k'} - \bar{Z}_{\cdot\cdot k})^2}{S\times B{:}W_{ZZ}}\right]}$$

$$SE(\bar{Y}_{i\cdot k\ adj} - \bar{Y}_{i\cdot k'\ adj}) = \sqrt{E_b\left[\frac{2}{r} + \frac{(\bar{Z}_{i\cdot k'} - \bar{Z}_{i\cdot k})^2}{S\times B{:}W_{ZZ}}\right]}$$

and, for $i \neq i'$

$$SE(\bar{Y}_{i\cdot k\ adj} - \bar{Y}_{i'\cdot k'\ adj}) = \left\{\frac{2[E_a+(s-1)E_b]}{rs} + \frac{(\bar{Z}_{i'\cdot\cdot} - \bar{Z}_{i\cdot\cdot})^2}{W\times B_{ZZ}} E_a + \frac{(\bar{Z}_{i'\cdot k'} - \bar{Z}_{i'\cdot\cdot} - \bar{Z}_{i\cdot k} + \bar{Z}_{i\cdot\cdot})^2}{S\times B{:}W_{ZZ}} E_b\right\}^{1/2} .$$

Variances and standard errors of adjusted means and differences amongst adjusted means for SP-3.

$$Var(\bar{Y}_{i\cdot k\ adj}) = (\sigma^2_\epsilon + \sigma^2_\delta)/r + (\sigma^2_\epsilon + s\sigma^2_\delta)\ \frac{(\bar{Z}_{i\cdot} - \bar{Z}_{\cdot\cdot})^2}{E(a)_{ZZ}}$$

$$Var(\bar{Y}_{i\cdot\cdot\ adj}) = (\sigma^2_\epsilon + s\sigma^2_\delta)\left[\frac{1}{sr} + \frac{(\bar{Z}_{i\cdot} - \bar{Z}_{\cdot\cdot})^2}{E(a)_{ZZ}}\right]$$

$$Var(\bar{Y}_{\cdot\cdot k\ adj}) = (\sigma^2_\epsilon + \sigma^2_\delta)/ar = Var(\bar{Y}_{\cdot\cdot k})$$

$$Var(\bar{Y}_{i\cdot\cdot\ adj} - \bar{Y}_{i'\cdot\cdot\ adj}) = (\sigma^2_\epsilon + s\sigma^2_\delta)\left[\frac{2}{sr} + \frac{(\bar{Z}_{i'\cdot} - \bar{Z}_{i\cdot})^2}{E(a)_{ZZ}}\right]$$

$$Var(\bar{Y}_{\cdot\cdot k\ adj} - \bar{Y}_{\cdot\cdot k'\ adj}) = 2\ \sigma^2_\epsilon/ar = Var(\bar{Y}_{\cdot\cdot k} - \bar{Y}_{\cdot\cdot k'})$$

$$Var(\bar{Y}_{i\cdot k\ adj} - \bar{Y}_{i\cdot k'\ adj}) = 2\ \sigma^2_\epsilon/r = Var(\bar{Y}_{i\cdot k} - \bar{Y}_{i\cdot k'})$$

and for $i \neq i'$,

$$\mathrm{Var}(\bar{Y}_{i\cdot k\ adj} - \bar{Y}_{i'\cdot k'\ adj}) = 2(\sigma_\epsilon^2 + \sigma_\delta^2)/r + (\sigma_\epsilon^2 + s\sigma_\delta^2)\ \frac{(\bar{Z}_{i'\cdot} - \bar{Z}_{i\cdot})^2}{E(a)_{ZZ}}$$

Estimates of the variance components σ_ϵ^2 and σ_δ^2 are given by E(b) and [E(a) - E(b)]/s, respectively. Hence, the desired standard errors are given by:

$$SE(\bar{Y}_{i\cdot\cdot\ adj} - \bar{Y}_{i'\cdot\cdot\ adj}) = \sqrt{E(a)\left[\frac{2}{sr} + \frac{(\bar{Z}_{i'\cdot} - \bar{Z}_{i\cdot})^2}{E(a)_{ZZ}}\right]}$$

$$SE(\bar{Y}_{\cdot\cdot k\ adj} - \bar{Y}_{\cdot\cdot k'adj}) = \sqrt{2E(b)/ar} = SE(\bar{Y}_{\cdot\cdot k} - \bar{Y}_{\cdot\cdot k'})$$

$$SE(\bar{Y}_{i\cdot k\ adj} - \bar{Y}_{\cdot\cdot k'\ adj}) = \sqrt{2E(b)/r} = SE(\bar{Y}_{i\cdot k} - \bar{Y}_{i\cdot k'})$$

and, for $i \neq i'$,

$$SE(\bar{Y}_{i\cdot k\ adj} - \bar{Y}_{i'\cdot k'}) = \sqrt{\frac{2[E(a) + (s-1)E(b)]}{sr} + \frac{(\bar{Z}_{i'\cdot} - \bar{Z}_{i\cdot})^2}{E(a)_{ZZ}}\ E(a)}$$

References

Federer, W.T. (1955) Experimental Design - Theory and Application. The MacMillan Co., New York, Chapter 16.

Federer, W.T., and Henderson, H.V. (1979), Covariance analysis of design experiments x statistical packages: An Update, Proc., Comp. Sci. an Stat. 12th Ann. Sym. on the Interface.

SAS User's Guide: Statistics, Version 5 Edition (1985a), SAS Institute Inc., Cary, NC: 956pp.

SAS User's Guide: Basics, Version 5 Edition (1985b), SAS Institute Inc., Cary, NC: 1290pp.

Searle, S.R., (1971), Linear Models, Wiley, New York, 532pp.

Searle, S.R., Hudson, G.F.S., and Federer, W.T. (1985), Annotated computer output for covariance-text, BU-780-M, Biometrics Unit Mimeo Ser., Cornell University, Ithaca, New York.

Winer, B.J., (1971), Statistical Principles in Experimental Design McGraw-Hill Book Company, New York, 907pp.

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
MODEL	11 SSRm	328.00000000	29.81818182
ERROR	12 SSE(b)	112.00000000	9.33333333
CORRECTED TOTAL	23 SSTm	440.00000000	

ρ_j = block effect
τ_i = whole plot effect
α_k = subplot effect

MODEL F = $3.19 = \frac{29.82}{9.33}$ PR > F = 0.0288

R-SQUARE	C.V.	ROOT MSE	Y MEAN = overall mean
0.745455	43.6436	3.05505046	7.00000000

NOTE: These data are balanced. Therefore, type I,II,III, and IV SS's are equal, so only type III SS's were required. Type I SS's are the cheapest.

<u>Wrong Test</u>
SAS computes all tests in this part of the table using SSE(b)=112.000. The appropriate test of whole plot effects uses SSE(a)=16.000 and must be requested using TEST command.

SOURCE	DF		TYPE III SS	F VALUE	PR > F
BLOCK	2	$R(\rho\|\mu,\tau,\alpha,\alpha\tau)$	48.00000000	2.57	0.1176
WHOLE (plot)	1	$R(\tau\|\mu,\rho,\alpha,\alpha\tau)$	24.00000000	2.57	0.1348
SUBPLOT	3	SSE(α)	156.00000000	5.57	0.0125
BLOCK*WHOLE	2	$R(\delta\|\mu,\rho,\tau,\alpha\tau)$	16.00000000	0.86	0.4488
WHOLE*SUBPLOT	3	$R(\alpha\tau\|\mu,\rho,\tau,\alpha)$	84.00000000	3.00	0.0728

TESTS OF HYPOTHESES USING THE TYPE III MS FOR BLOCK*WHOLE AS AN ERROR TERM

SOURCE	DF		TYPE III SS	F VALUE	PR > F
BLOCK	2	$R(\rho\|\mu,\tau,\alpha,\alpha\tau)$	48.00000000	$3.00 = \frac{48.00/2}{16.00/2} = 3.00$	0.2500
WHOLE (plot)	1	$R(\tau\|\mu,\rho,\alpha,\alpha\tau)$	24.00000000	3.00	0.2254

Results of TEST command

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

OBSERVATION	Y = OBSERVED VALUE	$\hat{Y}$=xb = PREDICTED VALUE	Y-$\hat{Y}$ = RESIDUAL
1	Y_{111} = 3.00000000	$\hat{Y}_{111}$ = 4.00000000	-1.00000000 = $Y_{111} - \hat{Y}_{111}$
2	4.00000000	7.00000000	-3.00000000
3	7.00000000	3.00000000	4.00000000
4	6.00000000	6.00000000	0.00000000
5	6.00000000	6.00000000	-0.00000000
6	10.00000000	9.00000000	1.00000000
7	1.00000000	5.00000000	-4.00000000
8	11.00000000	8.00000000	3.00000000
9	6.00000000	5.00000000	1.00000000
10	10.00000000	8.00000000	2.00000000
11	4.00000000	4.00000000	0.00000000
12	4.00000000	7.00000000	-3.00000000
13	3.00000000	4.00000000	-1.00000000
14	2.00000000	3.00000000	-1.00000000
15	1.00000000	1.00000000	-0.00000000
16	14.00000000	12.00000000	2.00000000
17	8.00000000	8.00000000	0.00000000
18	8.00000000	7.00000000	1.00000000
19	2.00000000	5.00000000	-3.00000000
20	18.00000000	16.00000000	2.00000000
21	10.00000000	9.00000000	1.00000000
22	8.00000000	8.00000000	-0.00000000
23	9.00000000	6.00000000	3.00000000
24	13.00000000	17.00000000	-4.00000000

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS = SSE(b)	112.00000000
SUM OF SQUARED RESIDUALS - ERROR SS	-0.00000000
FIRST ORDER AUTOCORRELATION	-0.31250000
DURBIN-WATSON D	2.47321429

First order auto correlation and Durbin-Watson D are tests used to detect time-correlated errors. Not applicable for these data. See for example Neter and Wasserman (1974).

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF		SUM OF SQUARES	MEAN SQUARE	F VALUE
MODEL	4	SSRm	88.00000000	22.00000000	99999.99
ERROR	1	SSE(a)	0.00000000	0.00000000	PR > F
CORRECTED TOTAL	5	SSTm	88.00000000		0.0001

NOTE: To get SS's for the whole plot from split plot data, it is necessary to use totals of each variety by block combination divided by $\sqrt{4}$ = $\sqrt{\text{no. of split plot treatments}}$. The following data are used for this run.

R-SQUARE	C.V.	ROOT MSE	Y MEAN
1.000000	0.0000	0.00000000	14.00000000 = $Y_{...}(\sqrt{4})$ = 7·2 = 14

SOURCE	DF		TYPE I SS	F VALUE	PR > F
BLOCK	2	$R(\rho\|\mu)$	48.00000000	.	.
WHOLE	1	$R(\tau\|\mu,\rho)$	24.00000000	.	.
Z	1	$R(\beta_1\|\mu,\rho,\tau)$	16.00000000	.	.

NOTE: These data are balanced. Therefore, Types II,III, and IV are all equal. Since SSE(a) = 0, F ratio is undefined.

SOURCE	DF		TYPE III SS	F VALUE	PR > F
BLOCK	2	$R(\rho\|\mu,\tau,\beta_1,\alpha,\alpha\tau)$	15.60000000	.	.
WHOLE	1	$R(\tau\|\mu,\rho,\beta_1,\alpha,\alpha\tau)$	3.42857143	.	.
Z	1	$R(\beta_1\|\mu,\tau,\rho,\alpha,\alpha\tau)$	16.00000000	.	.

ρ_j = block effect
τ_i = whole plot effect
α_k = subplot effect
β_1 = Z(whole plot covariate slope)

PARAMETER			ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT		μo =	-12.00000000 B	-99999.99	0.0001	0
BLOCK	1	ρo_1 =	6.00000000 B	99999.99	0.0001	0
	2		6.00000000 B	99999.99	0.0001	0
	3		0.00000000 B	.	.	.
WHOLE	1	τo_1 =	4.00000000 B	99999.99	0.0001	0
	2		0.00000000 B	.	.	.
Z		$\hat{\beta}_1$ =	4.00000000	99999.99	0.0001	0

use $\pm$ 99999.99
to express that F is infinite

$$\mu o = Y_{...}(\sqrt{4}) - \frac{1}{3}\sum_{j=1}^{3} \rho o_j - \frac{1}{2}\sum_{i=1}^{2} \tau o_i - \hat{\beta}_1 Z_{...}(\sqrt{4})$$

$$= 7\cdot(\sqrt{4}) - \frac{1}{3}(6+6+0) - \frac{1}{2}(4+0) - 4.0(2.5)(\sqrt{4}) = -12$$

$$\rho o_1 = Y_{.1.}(\sqrt{4}) - \mu o - \frac{1}{2}\Sigma\, \tau o_i - \hat{\beta}_1 Z_{.1.}(\sqrt{4}) = 5(2) - (-12) - \frac{1}{2}(4+0) - 4(1.75)(\sqrt{4}) = 6$$

$$\tau o_1 = Y_{1..}(\sqrt{4}) - \mu o - \frac{1}{3}\Sigma\, \rho o_j - \hat{\beta}_1 Z_{i..}(\sqrt{4}) = 6(\sqrt{4}) - (-12) - \frac{1}{3}(6+6+0) - 4(2)(\sqrt{4}) = 4$$

These individual estimates, with the exception of the covariate estimate, are not useful to the experimenter. They can be used together to compute predicted values (see next page). More meaningful estimates (adjusted means) are printed on SP-2 page 4. Except for $\hat{\beta}_1$ these estimates are not of interest. It is preferable to use the ESTIMATE statement as shown below.

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
WHOLE PLOT SLOPE	4.00000000	99999.99	0.0001	0

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

NOTE: THE X'X MATRIX HAS BEEN DEEMED SINGULAR AND A GENERALIZED INVERSE HAS BEEN EMPLOYED TO SOLVE THE NORMAL EQUATIONS. THE ABOVE ESTIMATES REPRESENT ONLY ONE OF MANY POSSIBLE SOLUTIONS TO THE NORMAL EQUATIONS. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED AND DO NOT ESTIMATE THE PARAMETER BUT ARE BLUE FOR SOME LINEAR COMBINATION OF PARAMETERS (OR ARE ZERO). THE EXPECTED VALUE OF THE BIASED ESTIMATORS MAY BE OBTAINED FROM THE GENERAL FORM OF ESTIMABLE FUNCTIONS. FOR THE BIASED ESTIMATORS, THE STD ERR IS THAT OF THE BIASED ESTIMATOR AND THE T VALUE TESTS HO: E(BIASED ESTIMATOR) = 0. ESTIMATES NOT FOLLOWED BY THE LETTER B ARE BLUE FOR THE PARAMETER.

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	$Y_{11.}/\sqrt{4}$ = 10.00000000	$\hat{Y}_{11.}/\sqrt{4}$ = 10.00000000	-0.00000000 $= \frac{Y_{11.}}{\sqrt{4}} - \frac{\hat{Y}_{11.}}{\sqrt{4}}$
2	14.00000000	14.00000000	-0.00000000
3	12.00000000	12.00000000	-0.00000000
4	10.00000000	10.00000000	-0.00000000
5	18.00000000	18.00000000	0.00000000
6	20.00000000	20.00000000	-0.00000000

Remember that $Y_{11.}$ for this run is the sum of block 1 trt 1 divided by $\sqrt{4}$.
i.e. $(3+4+7+6)/\sqrt{4} = 10$

SUM OF RESIDUALS	-0.00000000
SUM OF SQUARED RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS - ERROR SS	0.00000000
FIRST ORDER AUTOCORRELATION	0.00000000
DURBIN-WATSON D	0.00000000

$$\hat{Y}_{11.}/\sqrt{4} = \mu^o + \rho^o_1 + \tau^o_1 + \hat{\beta}_1(Z_{11.})(\sqrt{4})$$

$$= -12 + 6 + 4 + 4(\tfrac{6}{4})(2) = 10$$

where $Z_{11.} = (1 + 2 + 1 + 2)/4$

GENERAL LINEAR MODELS PROCEDURE

LEAST SQUARES MEANS = Adjusted Means

BLOCK	Y LSMEAN
1	$Y_{.1._{adj}}$ = 16.0000000
2	16.0000000
3	10.0000000

$$Y_{.j._{adj}} = 2\left[Y_{.j.} - \hat{\beta}_1(Z_{.j.} - Z_{...})\right]$$

e.g. $Y_{.1._{adj}} = 2\left[\frac{20+20}{8} - 4(1.75 - 2.5)\right] = 2(8) = 16$

WHOLE (plot)	Y LSMEAN
1	$Y_{1.._{adj}}$ = 16.0000000
2	12.0000000

$$Y_{i.._{adj}} = 2\left[Y_{i..} - \hat{\beta}_1(Z_{i..} - Z_{...})\right]$$

e.g. $Y_{1.._{adj}} = 2\left[6 - 4(2 - 2.5)\right] = 2(8) = 16$

NOTE: Since we used totals $/(\sqrt{4} = 2)$ as input data, these adjusted means need to be divided by 2..

CORRECT SPLIT PLOT ANALYSIS
GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE
MODEL	12	342.45000000	28.53750000	3.22
ERROR	11	97.55000000	8.86818182 = $\hat{\sigma}^2$	PR > F
CORRECTED TOTAL	23	440.00000000		0.0312

R-SQUARE	C.V. = $\frac{\hat{\sigma}}{Y_{...}}$ × 100%	ROOT MSE	Y MEAN
0.778295 = $\frac{SS(Model)}{SS(Total)}$ = $\frac{343.45}{440.00}$	42.5421 = $\frac{2.9779}{7.0000}$ × 100%	2.97794926 = $\sqrt{MS_{error}}$	7.00000000

SOURCE	DF		TYPE I SS	F VALUE	PR > F
BLOCK	2	$R(\rho\|\mu)$	48.00000000	2.71	0.1107
WHOLE (plot)	1	$R(\tau\|\mu,\rho)$	24.00000000	2.71	0.1282
BLOCK*WHOLE	2	SSE(a)unadj	16.00000000	0.90	0.4337
SUBPLOT	3	$R(\alpha\|\mu,\rho,\tau)$	156.00000000	5.86	0.0121
WHOLE*SUBPLOT	3	$R(\alpha\tau\|\mu,\rho,\tau,\alpha)$	84.00000000	3.16	0.0683
Z	1	$R(\beta_2\|\mu,\rho,\tau,\alpha,\alpha\tau)$	14.45000000	1.63	0.2281

NOTE: These data are balanced. Therefore Types II, III, and IV are equal.

These 5 SS's may not be useful to the investigator as they are for the data unadjusted for the covariate.

SOURCE	DF	TYPE III SS		F VALUE	PR > F
BLOCK	2	20.20862069		1.14	0.3551
WHOLE	1	6.10384615		0.69	0.4244
BLOCK*WHOLE	2	9.45000000		0.53	0.6014
SUBPLOT	3	84.24310345	$R(\alpha\|\mu,\rho,\tau,\alpha\tau,\beta_2)$	3.17	0.0678
WHOLE*SUBPLOT	3	37.47439024	$R(\alpha\tau\|\mu,\rho,\tau,\alpha,\beta_2)$	1.41	0.2923
Z	1	14.45000000	$R(\beta_2\|\mu,\rho,\tau,\alpha,\alpha\tau)$	1.63	0.2281

These 3 SS's should be ignored because they are adjusted with β_2 instead of β_1.

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE

CORRECT SPLIT PLOT ANALYSIS

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

PARAMETER		ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT		μ^o = 11.90000000 B	2.63	0.0232	4.51628367
BLOCK	1	ρ^o_1 = -3.30000000 B	-1.32	0.2122	2.49153111
	2	-0.15000000 B	-0.07	0.9471	2.20850628
	3	0.00000000 B	.	.	.
WHOLE	1	τ^o_1 = -7.02500000 B	-1.86	0.0902	3.78152657
	2	0.00000000 B	.	.	.
BLOCK*WHOLE	1 1	3.15000000 B	1.03	0.3241	3.05148995
	1 2	0.00000000 B	.	.	.
	2 1	$(\delta)^o_{21}$ = 1.57500000 B	0.53	0.6096	2.99650364
	2 2	0.00000000 B	.	.	.
	3 1	0.00000000 B	.	.	.
	3 2	0.00000000 B	.	.	.
SUBPLOT	1	α^o_1 = -6.30000000 B	-2.27	0.0441	2.77231989
	2	-5.60000000 B	-1.55	0.1488	3.60647566
	3	-9.30000000 B	-3.35	0.0064	2.77231989
	4	0.00000000 B	.	.	.
WHOLE*SUBPLOT	1 1	$(\alpha\tau)^o_{11}$ = 4.30000000 B	1.17	0.2682	3.68753017
	1 2	5.75000000 B	1.20	0.2549	4.78638378
	1 3	7.15000000 B	2.04	0.0659	3.50252074
	1 4	0.00000000 B	.	.	.
	2 1	0.00000000 B	.	.	.
	2 2	0.00000000 B	.	.	.
	2 3	0.00000000 B	.	.	.
	2 4	0.00000000 B	.	.	.
Z		$\hat{\beta}_2$ = 0.85000000	1.28	0.2281	0.66588970

NOTE: Except for $\hat{\beta}_2$, none of these estimates are of interest. It is preferable to use the ESTIMATE statement to obtain $\hat{\beta}_2$.

$$\alpha^o_1 = Y_{\cdot\cdot 1} - \mu^o - \frac{1}{2}\sum_{i=1}^{2}\tau^o_i - \frac{1}{3}\sum_{j=1}^{3}\rho^o_j - \frac{1}{6}\sum_{j=1}^{3}\sum_{i=1}^{2}\delta^o_{ij} - \frac{1}{2}\sum_{i=1}^{2}(\alpha\tau)^o_{ik} - \hat{\beta}_2 Z_{\cdot\cdot k}$$

$$= 6 - 11.9 - \frac{1}{2}(-7.025 + 0) - \frac{1}{3}(-3.3 - 0.15 + 0) - \frac{1}{6}(3.15 + 1.575 + 0 + 0 + 0 + 0) - \frac{1}{2}(4.30 + 0) - 0.85(2.5)$$

$$= -6.3$$

$$(\tau\alpha)^o_{i\cdot k} = Y_{i\cdot k} - \mu^o - \tau^o_i - \alpha^o_k - \frac{1}{3}\sum_{j=1}^{3}\rho^o_j - \frac{1}{3}\sum_{j=1}^{3}\delta_{ij} - \hat{\beta}_2 Z_{i\cdot k}$$

$$(\tau\alpha)^o_{1\cdot 1} = \left(\frac{15}{3}\right) - 11.9 + 7.025 + 6.3 - \frac{1}{3}(-3.3 - 0.15) - \frac{1}{3}(3.15 + 1.575 + 0) - 0.85\left(\frac{6}{3}\right)$$

$$= 4.3$$

CORRECT SPLIT PLOT ANALYSIS
GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	$Y_{111} = 3.00000000$	$\hat{Y}_{111} = 3.57500000$	$Y_{111} - \hat{Y}_{111} = -0.57500000$
2	4.00000000	6.57500000	-2.57500000
24	13.00000000	17.85000000	-4.85000000

for split plot

$$\hat{Y}_{111} = \mu^{o} + \rho^{o}_{1} + \tau^{o}_{1} + \delta^{o}_{11} + \alpha^{o}_{1} + (\alpha\tau)^{o}_{11} + \hat{\beta}_2 Z_{111}$$

$$= 11.9 + (-3.3) + (-7.025) + 3.15 + (-6.3) + 4.3 + (.85)(1)$$

$$= 3.575$$

SUM OF RESIDUALS	-0.00000000
SUM OF SQUARED RESIDUALS	97.55000000
SUM OF SQUARED RESIDUALS - ERROR SS	-0.00000000
FIRST ORDER AUTOCORRELATION	-0.40431830
DURBIN-WATSON D	2.56411456

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
SUBPLOT SLOPE	0.85000000	1.28	0.2281	0.66588970

This is the result from the ESTIMATE statement used to find $\hat{\beta}_2$.

CORRECT SPLIT PLOT ANALYSIS
GENERAL LINEAR MODELS PROCEDURE

LEAST SQUARES MEANS

SUBPLOT	Y LSMEAN
1	6.0000000
2 $Y_{\cdot\cdot 2_{adj}}$ =	7.4250000
3	4.4250000
4	10.1500000

$Y_{\cdot\cdot 2_{adj}} = Y_{\cdot\cdot 2} - \hat{\beta}_2(Z_{\cdot\cdot 2} - Z_{\cdot\cdot\cdot}) = 7 - 0.85(2 - 2.5) = 7.425$

WHOLE	SUBPLOT	Y LSMEAN
1	1	5.4250000
1	2	7.5750000
1	3	5.2750000
1	4	7.4250000
2	1	6.5750000
2	2	7.2750000
2	3	3.5750000
2	4	12.8750000

$12.8750000 = Y_{2\cdot 4} - \hat{\beta}_2(Z_{2\cdot 4} - Z_{\cdot\cdot\cdot}) = 15 - 0.85(5 - 2.5) = 12.875$

These adjusted means for a split plot treatment are wrong because they are not adjusted for the whole plot regression as well as the split plot regression. The appropriate adjustment would be

$$Y_{i\cdot k} - \hat{\beta}_1(Z_{i\cdot\cdot} - Z_{\cdot\cdot\cdot}) - \hat{\beta}_2(Z_{i\cdot k} - Z_{i\cdot\cdot}) = Y_{i\cdot k\ adj}$$

These can easily be computed by combining the output from both procedural cells.

$$\bar{y}_{1\cdot 1adj} = 5 - 4(2 - 2.5) - 0.85(2 - 2) = 7.$$

SP-3: SPLIT PLOTS WITH WHOLE PLOT ARRANGED IN CRD WITH A
COVARIATE CONSTANT OVER SPLIT PLOTS
WHOLE PLOT ANALYSIS
12:47 THURSDAY, OCTOBER 2, 1986

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: MY

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
MODEL	2 SSRm	234.63930251	117.31965125
ERROR	5 SSE(a)	61.29819749	12.25963950
CORRECTED TOTAL	7 STTm	295.93750000	

Note: To get the SS's for the whole plot from split plot data, it is necessary to use totals of each subject divided by $\sqrt{2}$. Those data follow and are used for this run.

SUBJECT	A1 Z	A1 Y	SUBJECT	A2 Z	A2 Y
1	4.2	12.7	5	1.4	17.7
2	7.1	19.1	6	11.3	31.8
3	11.3	24.0	7	1.4	24.7
4	2.8	12.7	8	2.8	17.7

MODEL F = 9.57 PR > F = 0.0195

R-SQUARE	C.V.	ROOT MSE	MY MEAN = Y...($\sqrt{2}$)
0.792868	17.4509	3.50137680	20.06415492

SOURCE	DF		TYPE I SS	F VALUE	PR > F
Z	1	$R(\beta_1 \| \mu)$	190.14770093	15.51	0.0110
A	1	$R(\tau \| \mu, \beta_1)$	44.49160158	3.63	0.1151

τ_i = A effect (whole plot)
α_k = B effect (split plot)
β_1 = Z (covariate slope)

SOURCE	DF		TYPE III SS	F VALUE	PR > F
Z	1	$R(\beta_1 \| \mu, \tau)$	166.57680251	13.59	0.0142
A	1	$R(\tau \| \mu, \beta_1)$	44.49160158	3.63	0.1151

Note: These data are balanced. Therefore Type II,III, are IV SS's are all equal.

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: MY

PARAMETER		ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT		μ^0= 15.39342646 B	5.70	0.0023	2.70221754
Z		$\hat{\beta}_1$= 1.02194357	3.69	0.0142	0.27724167
A	1	τ_1^0= -4.74969610 B	-1.91	0.1151	2.49324900
	2	0.00000000 B	.	.	.

$$\mu^0 = Y_{...}(\sqrt{2}) - \frac{1}{2}\sum_{i=1}^{2}\tau_i^0 - \hat{\beta}_1 Z_{...}(\sqrt{2})$$

$$= 14.1875(\sqrt{2}) - \frac{1}{2}(-4.749696 + 0) - 1.02194(4.875)(\sqrt{2})$$

$$= 15.3934$$

$$\tau_1^0 = Y_{1..}(\sqrt{2}) - \mu^0 - \hat{\beta}_1 Z_{1..}(\sqrt{2})$$

$$= 12.125(\sqrt{2}) - 15.393426 - 1.02194\ (4.5)(\sqrt{2})$$

where $Y_{1..} = \frac{18+27+34+18}{8} = 12.125$

$$Z_{1..} = \frac{3+5+8+2}{4} = 4.5$$

DEPENDENT VARIABLE: MY

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL for whole plot
1	$\bar{y}_{11.}(\sqrt{2})$ = 12.72792206	$\hat{y}_{11.}(\sqrt{2})$ = 14.97946975	$(\bar{y}_{11.}-\hat{y}_{11.})(\sqrt{2})$ = -2.25154769
2	19.09188309	17.86996267	1.22192042
3	24.04163056	22.20570206	1.83592850
4	12.72792206	13.53422329	-0.80630123
5	17.67766953	16.83867293	0.83899660
6	31.81980515	26.95539816	4.86440699
7	24.74873734	29.84589108	-5.09715374
8	17.67766953	18.28391939	-0.60624986

$$y_{11.}(\sqrt{2}) = \mu^0 + \tau_i^0 + \hat{\beta}_1 Z_{11.}(\sqrt{2})$$
$$= 15.393 - 4.750 + 1.445(3)$$
$$= 14.979$$

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	61.29819749
SUM OF SQUARED RESIDUALS - ERROR SS	-0.00000000
FIRST ORDER AUTOCORRELATION	-0.33097074
DURBIN-WATSON D	2.57324383

LEAST SQUARES MEANS

A	MY (whole plot) LSMEAN
1	17.6893069 $= (\bar{y}_{i..} - \hat{\beta}_1(Z_{i..} - Z_{...}))\sqrt{2}$ = correct $Y_{1..adj}(\sqrt{2})$
2	22.4390030

Note: Divide by $\sqrt{2}$ to get correct adjusted mean. The correct adjusted means would have resulted if the average for each subject was used as data but the correct ANOVA would not.

$$\text{correct } Y_{1..adj} = (Y_{1..} - \hat{\beta}_1(Z_{i..} - Z_{...}))\sqrt{2}/\sqrt{2}$$
$$= [12.125 - 1.022(4.5) - 4.875)]\sqrt{2}/\sqrt{2}$$
$$= 17.6893/\sqrt{2}$$
$$= 12.51$$

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR ESTIMATE
REGR SLOPE	1.02194357	3.69	0.0142	0.27724167

LEAST SQUARES MEANS

A	MY LSMEAN	STD ERR LSMEAN	PROB > |T| H0:LSMEAN=0
1	17.6893069	1.7568516	0.0002
2	22.4390030	1.7568516	0.0001

SP-3: SPLIT PLOTS WITH WHOLE PLOT ARRANGED IN CRD WITH A
COVARIATE CONSTANT OVER SPLIT PLOTS
SPLIT PLOT ANALYSIS
12:47 THURSDAY, OCTOBER 2, 1986

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF		SUM OF SQUARES	MEAN SQUARE
MODEL	9	SSRm	382.06250000	42.45138889
ERROR	6	SSE(b)	6.37500000	1.06250000
CORRECTED TOTAL	15	SSTm	388.43750000	

MODEL F = 39.95 PR > F = 0.0001

R-SQUARE	C.V.	ROOT MSE	Y MEAN
0.983588	7.2654	1.03077641	14.18750000

SOURCE	DF		TYPE III SS	F VALUE	PR > F
A (whole plot)	1	$R(\tau\|\mu,\alpha,\alpha\tau)$	68.06250000	64.06	0.0002
SUB (A)	6	SSEa(unadjusted)	227.87500000	35.75	0.0002
B (split plot)	1	$R(\alpha\|\mu,\tau,\alpha\tau)$	85.56250000	80.53	0.0001
A*B	1	$R(\alpha\tau\|\mu,\tau,\alpha)$	0.56250000	0.53	0.4943

Because these data are balanced and there is no split plot covariate, all 4 types of SS's are equal. Type I SS's would be cheapest to compute.

These SS's are not used since they are not corrected by β_1 .

These SS's are reported in the ANOVA table.

SP-3: SPLIT PLOTS WITH WHOLE PLOT ARRANGED IN CRD WITH A
COVARIATE CONSTANT OVER SPLIT PLOTS
SPLIT PLOT ANALYSIS
12:47 THURSDAY, OCTOBER 2, 1986

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

OBSERVATION		OBSERVED VALUE	PREDICTED VALUE	RESIDUAL	SPLIT PLOT
1	$Y_{111}=$	10.00000000	$\hat{Y}_{111}$=11.12500000	$Y_{111}-\hat{Y}_{111}$=-1.12500000	$\hat{Y}_{111}=\mu^o + \tau_1^o + \delta_{111}^o + \alpha_1^o + \tau\alpha_{11}^o$
2		8.00000000	6.87500000	1.12500000	=10.000 - 3.125 + 0 + 5.000 - .75
:		:	:	:	
16		10.00000000	10.00000000	0.00000000	

SUM OF RESIDUALS	-0.00000000
SUM OF SQUARED RESIDUALS	6.37500000
SUM OF SQUARED RESIDUALS - ERROR SS	0.00000000
FIRST ORDER AUTOCORRELATION	-0.64460784
DURBIN-WATSON D	3.09068627

LEAST SQUARES MEANS = unadjusted means

B	Y LSMEAN
1	16.5000000 = $Y_{\cdot\cdot k}$
2	11.8750000

A	B	Y LSMEAN
1	1	14.2500000
1	2	10.0000000 = $Y_{i\cdot k}$
2	1	18.7500000
2	2	13.750000

SAS does not give the adjusted means for $Y_{ij\cdot}$. These can be computed by hand using the following formula:

$$Y_{i\cdot k}\text{adj} = Y_{i\cdot k} - \hat{\beta}_1(Z_{i\cdot\cdot} - Z_{\cdot\cdot\cdot})$$

Procedural call for SP-2A

```
DATA ONE;
INPUT EU BLOCK WHOLE SUBPLOT Z ZTOTAL Y;  CARDS;

   { data }

TITLE 'SPLIT PLOT HYPOTHETICAL DATA: COVARIATE ADDED';
PROC PRINT; VAR EU BLOCK WHOLE SUBPLOT Z ZTOTAL Y;

PROC GLM;  CLASS WHOLE BLOCK SUBPLOT;
   MODEL Y = BLOCK WHOLE ZTOTAL BLOCK*WHOLE
             SUBPLOT SUBPLOT*WHOLE Z / SS1 SS3 P;
   RANDOM BLOCK BLOCK*WHOLE;

   TEST H=ZTOTAL E=BLOCK*WHOLE / HTYPE=1 ETYPE=1;
   TEST H=WHOLE  E=BLOCK*WHOLE / HTYPE=3 ETYPE=3;

   ESTIMATE 'SUBPLOT SLOPE' Z 1;
   LSMEANS SUBPLOT / STDERR PDIFF;
```

⇒ **The ordering in the MODEL statement *is* important. RANDOM option prints expected mean squares for different Types of SS's.**

⇒ **TEST $H_o: \beta_1 = 0$ and whole plot main effects using the appropriate Type SS's for hypothesis SS's and error SS's.**

The LSMEANS statement gives correctly adjusted subplot means, however the reported standard errors are incorrect.

The ESTIMATE statement provides the estimate of the subplot slope coefficient β_2.

Unfortunately, the whole plot slope coeffient β_1 may not be estimated as easily.

SPLIT PLOT HYPOTHETICAL DATA: COVARIATE ADDED

OBS	EU	BLOCK	WHOLE	SUBPLOT	Z	ZTOTAL	Y
1	1	1	1	1	1	6	3
2	1	1	1	2	2	6	4
3	1	1	1	3	1	6	7
4	1	1	1	4	2	6	6
5	2	2	1	1	2	8	6
6	2	2	1	2	2	8	10
7	2	2	1	3	0	8	1
8	2	2	1	4	4	8	11
9	3	3	1	1	3	10	6
10	3	3	1	2	5	10	10
11	3	3	1	3	2	10	4
12	3	3	1	4	0	10	4
13	4	1	2	1	2	8	3
14	4	1	2	2	0	8	2
15	4	1	2	3	2	8	1
16	4	1	2	4	4	8	14
17	5	2	2	1	4	12	8
18	5	2	2	2	1	12	8
19	5	2	2	3	3	12	2
20	5	2	2	4	4	12	18
21	6	3	2	1	3	16	10
22	6	3	2	2	2	16	8
23	6	3	2	3	4	16	9
24	6	3	2	4	7	16	13

⇒ **Begin output from PROC PRINT**

ZTOTAL are the $Z_{ij.}$'s
EU is an indicator for the whole plot experimental units.

```
GENERAL LINEAR MODELS PROCEDURE
   CLASS LEVEL INFORMATION
 CLASS      LEVELS    VALUES
 WHOLE         2       1 2
 BLOCK         3       1 2 3
 SUBPLOT       4       1 2 3 4
NUMBER OF OBSERVATIONS IN DATA SET = 24
```

⇒ Begin output from PROC GLM

```
SOURCE         TYPE I EXPECTED MEAN SQUARE
BLOCK          VAR(ERROR) + 4 VAR(WHOLE*BLOCK) + 8 VAR(BLOCK) + Q(ZTOTAL,Z)
WHOLE          VAR(ERROR) + 4 VAR(WHOLE*BLOCK)
               + Q(WHOLE,ZTOTAL,WHOLE*SUBPLOT,Z)
ZTOTAL         VAR(ERROR) + 4 VAR(WHOLE*BLOCK) + Q(ZTOTAL,Z)
WHOLE*BLOCK    VAR(ERROR) + 4 VAR(WHOLE*BLOCK)
SUBPLOT        VAR(ERROR) + Q(SUBPLOT,WHOLE*SUBPLOT,Z)
WHOLE*SUBPLOT  VAR(ERROR) + Q(WHOLE*SUBPLOT,Z)
Z              VAR(ERROR) + Q(Z)
```

Note that these expected mean squares indicate the appropriate SS's to be used in the reported ANOVA table as well as indicating correct error terms for use in computing F-statistics.

```
SOURCE         TYPE III EXPECTED MEAN SQUARE
BLOCK          VAR(ERROR) + 4 VAR(WHOLE*BLOCK) + 4.4 VAR(BLOCK)
WHOLE          VAR(ERROR) + 4 VAR(WHOLE*BLOCK) + Q(WHOLE,WHOLE*SUBPLOT)
ZTOTAL         0
WHOLE*BLOCK    VAR(ERROR) + 4 VAR(WHOLE*BLOCK)
SUBPLOT        VAR(ERROR) + Q(SUBPLOT,WHOLE*SUBPLOT)
WHOLE*SUBPLOT  VAR(ERROR) + Q(WHOLE*SUBPLOT)
Z              VAR(ERROR) + Q(Z)
```

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE
MODEL	12	342.45000000	28.53750000	3.22
ERROR	11	97.55000000	8.86818182	PR > F
CORRECTED TOTAL	23	440.00000000		0.0312

⇒ Note that 8.8682 = Error(b), the subplot error term

R-SQUARE	C.V.	ROOT MSE	Y MEAN
0.778295	42.5421	2.97794926	7.00000000

SOURCE		DF	TYPE I SS	F VALUE	PR > F
BLOCK	= $R(\rho \mid \mu,\tau)$	2	**48.00000000**	2.71	0.1107
WHOLE	= $R(\tau \mid \mu,\rho)$	1	24.00000000	2.71	0.1282
ZTOTAL	= $R(\delta \mid \mu,\beta_1,\rho,\tau)$	1	**16.00000000**	1.80	0.2063
WHOLE*BLOCK	= $R(\delta \mid \mu,\rho,\tau,\beta_1)$	1	**0.00000000**	0.00	1.0000
SUBPLOT	= $R(\alpha \mid \mu,\rho,\tau)$	3	156.00000000	5.86	0.0121
WHOLE*SUBPLOT	= $R(\alpha\tau \mid \mu,\rho,\tau,\alpha)$	3	84.00000000	3.16	0.0683
Z	= $R(\beta_2 \mid \mu,\rho,\tau,\alpha,\alpha\tau)$	1	**14.45000000**	1.63	0.2281

<u>NOTE</u>: The boldface SS's are those that appear in the correct ANOVA table, as verified by the expected mean squares.

SOURCE		DF	TYPE III SS	F VALUE	PR > F
BLOCK		2	15.60000000	0.88	0.4422
WHOLE	= $R(\tau \mid \mu,\rho,\beta_1)$	1	**3.42857143**	0.39	0.5468
ZTOTAL		0	0.00000000	.	.
WHOLE*BLOCK	= $R(\delta \mid \mu,\rho,\tau,\beta_1)$	1	0.00000000	0.00	1.0000
SUBPLOT	= $R(\alpha \mid \mu,\rho,\tau,\alpha\tau,\beta_2)$	3	**84.24310345**	3.17	0.0678
WHOLE*SUBPLOT	= $R(\alpha\tau \mid \mu,\rho,\tau,\alpha,\beta_2)$	3	**37.47439024**	1.41	0.2923
Z	= $R(\beta_2 \mid \mu,\rho,\tau,\alpha,\alpha\tau)$	1	**14.45000000**	1.63	0.2281

TESTS OF HYPOTHESES USING THE TYPE I MS FOR WHOLE*BLOCK AS AN ERROR TERM

SOURCE	DF	TYPE I SS	F VALUE	PR > F
ZTOTAL	1	16.00000000	.	.

TESTS OF HYPOTHESES USING THE TYPE III MS FOR WHOLE*BLOCK AS AN ERROR TERM

SOURCE	DF	TYPE III SS	F VALUE	PR > F
WHOLE	1	3.42857143	.	.

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
SUBPLOT SLOPE	0.85000000	1.28	0.2281	0.66588970

⇒ **This is the result of the ESTIMATE statement to estimate the subplot regression coefficient.**

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	3.00000000	3.57500000	-0.57500000
2	4.00000000	6.57500000	-2.57500000
:	:	:	:
:	:	:	:
23	9.00000000	6.00000000	3.00000000
24	13.00000000	17.85000000	-4.85000000

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	97.55000000
SUM OF SQUARED RESIDUALS - ERROR SS	-0.00000000

LEAST SQUARES MEANS

SUBPLOT	Y LSMEAN	STD ERR LSMEAN	PROB > |T| H0:LSMEAN=0	LSMEAN NUMBER
1	6.0000000	1.2157427	0.0004	1
2	7.4250000	1.2605089	0.0001	2
3	4.4250000	1.2605089	0.0049	3
4	10.1500000	1.3861599	0.0001	4

These standard errors are *not* correct as they include only the subplot error term (see discussion regarding standard errors of various means on pages 14 to 16).

PROB > |T| H0: LSMEAN(I)=LSMEAN(J)

I/J	1	2	3	4
1	.	0.4331	0.3877	0.0458
2	0.4331	.	0.1088	0.1979
3	0.3877	0.1088	.	0.0150
4	0.0458	0.1979	0.0150	.

These p-values are correct for the pairwise comparison of subplot means since these differences depend only upon the subplot error term.

Additional procedural call for SP-2A

{ Same input as previous call}

```
PROC GLM;
     CLASS WHOLE BLOCK SUBPLOT;
     MODEL Y Z = BLOCK WHOLE BLOCK*WHOLE SUBPLOT WHOLE*SUBPLOT / SS1;
     MEANS WHOLE BLOCK WHOLE*BLOCK SUBPLOT WHOLE*SUBPLOT;
     MANOVA H=WHOLE E=BLOCK*WHOLE / PRINTE;
     MANOVA H=SUBPLOT / PRINTE;
```

The MEANS statement gives the unadjusted means of both the response Y and the covariate Z.

⇒ This call to GLM is *unecessary* if only the correct ANOVA table is desired. However, this call does give the information necessary to estimate any adjusted mean, the slope estimates, and any standard errors required for estimation.

The MANOVA statements give the B×W$_{YZ}$ and S×B:W$_{YZ}$ terms which may be used to calculate the whole plot and subplot slope estimates.

```
                    GENERAL LINEAR MODELS PROCEDURE

                       CLASS LEVEL INFORMATION
                     CLASS      LEVELS     VALUES
                     WHOLE         2       1 2
                     BLOCK         3       1 2 3
                     SUBPLOT       4       1 2 3 4

                 NUMBER OF OBSERVATIONS IN DATA SET = 24
```

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE
MODEL	11	328.00000000	29.81818182	3.19
ERROR	12	112.00000000	9.33333333	PR > F
CORRECTED TOTAL	23	440.00000000		0.0288

R-SQUARE	C.V.	ROOT MSE	Y MEAN
0.745455	43.6436	3.05505046	7.00000000

SOURCE	DF	TYPE I SS	F VALUE	PR > F
BLOCK	2	48.00000000	2.57	0.1176
WHOLE	1	24.00000000	2.57	0.1348
WHOLE*BLOCK	2	16.00000000	0.86	0.4488
SUBPLOT	3	156.00000000	5.57	0.0125
WHOLE*SUBPLOT	3	84.00000000	3.00	0.0728

⇒ **The same SS as in SP-1**

DEPENDENT VARIABLE: Z

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE
MODEL	11	46.00000000	4.18181818	2.51
ERROR	12	20.00000000	1.66666667	PR > F
CORRECTED TOTAL	23	66.00000000		0.0645

R-SQUARE	C.V.	ROOT MSE	Z MEAN
0.696970	51.6398	1.29099445	2.50000000

SOURCE	DF	TYPE I SS	F VALUE	PR > F
BLOCK	2	9.00000000	2.70	0.1076
WHOLE	1	6.00000000	3.60	0.0821
WHOLE*BLOCK	2	1.00000000	0.30	0.7462
SUBPLOT	3	9.00000000	1.80	0.2008
WHOLE*SUBPLOT	3	21.00000000	4.20	0.0301

⇒ **The same as in the table of sum of squares and cross products.**

MEANS

WHOLE	N	Y	Z
1	12	6.00000000	2.00000000
2	12	8.00000000	3.00000000

⇒ **These are the unadjusted means that are needed to compute adjusted means.**

SUBPLOT	N	Y	Z
1	6	6.0000000	2.50000000
2	6	7.0000000	2.00000000
3	6	4.0000000	2.00000000
4	6	11.0000000	3.50000000

WHOLE	SUBPLOT	N	Y	Z
1	1	3	5.0000000	2.00000000
1	2	3	8.0000000	3.00000000
1	3	3	4.0000000	1.00000000
1	4	3	7.0000000	2.00000000
2	1	3	7.0000000	3.00000000
2	2	3	6.0000000	1.00000000
2	3	3	4.0000000	3.00000000
2	4	3	15.0000000	5.00000000

E = TYPE I SS&CP MATRIX FOR: WHOLE*BLOCK

DF=2	Y	Z
Y	16.00000000 = $B{\times}W_{YY}$	4.00000000 = $B{\times}W_{YZ}$
Z	4.00000000	1.00000000 = $B{\times}W_{ZZ}$

E = ERROR SS&CP MATRIX

DF=12	Y	Z
Y	112.00000000 = $S{\times}B{:}W_{YY}$	17.00000000 = $S{\times}B{:}W_{YZ}$
Z	17.00000000	20.00000000 = $S{\times}B{:}W_{ZZ}$

Procedural call for SP-3A

```
DATA SP3;
INPUT SUBJECT A B Y Z; CARDS;
```

{ data }

```
PROC PRINT; VAR SUBJECT A B Y Z;
TITLE 'WHOLE-PLOTS IN CRD AND COVARIATE MEASURED ON WHOLE-PLOTS';
TITLE2 'SP-3 DATA FROM WINER, 1971, P.803.';

PROC GLM; CLASS SUBJECT A B;
     MODEL  Y = A Z SUBJECT(A) B A*B / SS1 SS3 P;
     RANDOM SUBJECT(A);

     TEST H=Z E=SUBJECT(A) / ETYPE=1 HTYPE=1;
     TEST H=A E=SUBJECT(A) / ETYPE=3 HTYPE=3;
     LSMEANS B / STDERR PDIFF;
```

⇒ **This call to GLM produces the correct ANOVA table for SP-3. Expected mean squares are printed with the RANDOM option and the TEST statement computes F-tests of H_o:β_1=0 and adjusted whole plot effects.**

```
PROC GLM; CLASS SUBJECT A B;
     MODEL  Y Z = A SUBJECT(A) B A*B / SS1;
     MEANS A B A*B;
     MANOVA H=A E=SUBJECT(A) / PRINTE;
```

⇒ **This second call to GLM is unecessary if only the correct ANOVA table is desired. However, it does give the correct table of sums of squares and cross products among Y and Z for SP-3 by using the MANOVA statement. These may be used to correctly estimate the whole plot slope coefficient. The MEANS option prints appropriate means for Y (unadjusted) and the covariate Z. Thus, sufficient information is given to calculate any adjusted means as well as appropriate standard errors.**

```
WHOLE-PLOTS IN CRD AND COVARIATE MEASURED ON WHOLE-PLOTS
          SP-3 DATA FROM WINER, 1971, P.803.

OBS   SUBJECT   A   B    Y    Z
  1      1      1   1   10    3
  2      2      1   1   15    5
  3      3      1   1   20    8
  :      :      :   :    :    :
  :      :      :   :    :    :
 15      7      2   2   15   10
 16      8      2   2   10    2
```

```
GENERAL LINEAR MODELS PROCEDURE
   CLASS LEVEL INFORMATION
  CLASS      LEVELS    VALUES
  SUBJECT       8      1 2 3 4 5 6 7 8
  A             2      1 2
  B             2      1 2
NUMBER OF OBSERVATIONS IN DATA SET = 16
```

⇒ **Begin output from first call to GLM**

```
SOURCE              TYPE I EXPECTED MEAN SQUARE
A                   VAR(ERROR) + 2 VAR(SUBJECT(A)) + Q(A,Z,A*B)
Z                   VAR(ERROR) + 2 VAR(SUBJECT(A)) + Q(Z)
SUBJECT(A)          VAR(ERROR) + 2 VAR(SUBJECT(A))
B                   VAR(ERROR) + Q(B,A*B)
A*B                 VAR(ERROR) + Q(A*B)
```

⇒ **The expected mean squares indicate the appropriate SS's to use for constructing F-tests.**

```
SOURCE              TYPE III EXPECTED MEAN SQUARE
A                   VAR(ERROR) + 2 VAR(SUBJECT(A)) + Q(A,A*B)
Z                   0
SUBJECT(A)          VAR(ERROR) + 2 VAR(SUBJECT(A))
B                   VAR(ERROR) + Q(B,A*B)
A*B                 VAR(ERROR) + Q(A*B)
```

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE
MODEL	9	382.06250000	42.45138889	39.95
ERROR	6	6.37500000	1.06250000	PR > F
CORRECTED TOTAL	15	388.43750000		0.0001

$= R(\tau,\alpha,\alpha\tau,\beta_1,\delta|\mu)$

Note that 1.0625 = Error(b)

R-SQUARE	C.V.	ROOT MSE	Y MEAN
0.983588	7.2654	1.03077641	14.18750000

SOURCE		DF	TYPE I SS	F VALUE	PR > F
A	$= R(\tau \mid \mu)$	1	68.06250000	64.06	0.0002
Z	$= R(\beta_1 \mid \mu,\tau)$	1	**166.57680251**	156.78	0.0001
SUBJECT(A)	$= R(\delta \mid \mu,\beta_1,\tau)$	5	**61.29819749**	11.54	0.0049
B	$= R(\alpha \mid \mu,\tau,\alpha\tau)$	1	**85.56250000**	80.53	0.0001
A*B	$= R(\alpha\tau \mid \mu,\tau,\alpha)$	1	**0.56250000**	0.53	0.4943

⇒ The boldface SS's correspond to those found in the correct ANOVA table.

SOURCE		DF	TYPE III SS	F VALUE	PR > F
A	$= R(\tau \mid \mu,\beta_1)$	1	**44.49160158**	41.87	0.0006
Z	$= R(\beta_1 \mid \mu,\tau,\delta)$	0	0.00000000	.	.
SUBJECT(A)	$= R(\delta \mid \mu,\beta_1,\tau)$	5	61.29819749	11.54	0.0049
B	$= R(\alpha \mid \mu,\tau,\alpha\tau)$	1	85.56250000	80.53	0.0001
A*B	$= R(\alpha\tau \mid \mu,\tau,\alpha)$	1	0.56250000	0.53	0.4943

Note that Error(a) $= R(\delta \mid \mu,\beta_1,\tau)/5 = 12.2596$

TESTS OF HYPOTHESES USING THE TYPE I MS FOR SUBJECT(A) AS AN ERROR TERM

SOURCE	DF	TYPE I SS	F VALUE	PR > F
Z	1	166.57680251	13.59	0.0142

TESTS OF HYPOTHESES USING THE TYPE III MS FOR SUBJECT(A) AS AN ERROR TERM

SOURCE	DF	TYPE III SS	F VALUE	PR > F
A	1	44.49160158	3.63	0.1151

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	10.00000000	11.12500000	-1.12500000
2	15.00000000	15.62500000	-0.62500000
3	20.00000000	19.12500000	0.87500000
:	:	:	:
:	:	:	:
15	15.00000000	15.00000000	0.00000000
16	10.00000000	10.00000000	-0.00000000

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	6.37500000
SUM OF SQUARED RESIDUALS - ERROR SS	-0.00000000

LEAST SQUARES MEANS

B	Y LSMEAN	STD ERR LSMEAN	PROB > |T| H0:LSMEAN=0	PROB > |T| H0: LSMEAN1=LSMEAN2
1	16.5000000	0.3644345	0.0001	0.0001
2	11.8750000	0.3644345	0.0001	

⇒ **These are the correct subplot means and correct p-value for H_o: LSMEAN1 = LSMEAN2, but the standard errors are incorrect (see discussion).**

GENERAL LINEAR MODELS PROCEDURE

⇒ **Begin output for the second call to GLM**

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
SUBJECT	8	1 2 3 4 5 6 7 8
A	2	1 2
B	2	1 2

NUMBER OF OBSERVATIONS IN DATA SET = 16

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE
MODEL	9	382.06250000	42.45138889	39.95
ERROR	6	6.37500000	1.06250000	PR > F
CORRECTED TOTAL	15	388.43750000		0.0001

R-SQUARE	C.V.	ROOT MSE	Y MEAN	
0.983588	7.2654	1.03077641	14.18750000	$= \bar{Y}_{\cdot\cdot}$

SOURCE	DF	TYPE I SS	F VALUE	PR > F
A	1	68.06250000	64.06	0.0002
SUBJECT(A)	6	227.87500000	35.75	0.0002
B	1	85.56250000	80.53	0.0001
A*B	1	0.56250000	0.53	0.4943

⇒ This is the ANOVA when the covariate Z is omitted. All Types of SS's will be the same since the data are balanced—Type I SS's are used as they are cheapest.

DEPENDENT VARIABLE: Z

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE
MODEL	9	161.75000000	17.97222222	99999.99
ERROR	6	0.00000000	0.00000000	PR > F
CORRECTED TOTAL	15	161.75000000		0.0001

The ERROR is zero because the covariate is constant over subplots.

R-SQUARE	C.V.	ROOT MSE	Z MEAN	
1.000000	0.0000	0.00000000	4.87500000	$= \bar{Z}_{\cdot\cdot}$

SOURCE	DF	TYPE I SS	F VALUE	PR > F
A	1	2.25000000	.	.
SUBJECT(A)	6	159.50000000	.	.
B	1	0.00000000	.	.
A*B	1	0.00000000	.	.

$= E(a)_{ZZ}$, needed for calculating the estimate of β_1.

These SS's are zero since the covariate is constant over subplots.

MEANS

A	N	Y	Z
1	8	12.1250000	4.50000000
2	8	16.2500000	5.25000000

These are the unadjusted Y means and the means of the covariate Z.

B	N	Y	Z
1	8	16.5000000	4.87500000
2	8	11.8750000	4.87500000

A	B	N	Y	Z
1	1	4	14.2500000	4.50000000
1	2	4	10.0000000	4.50000000
2	1	4	18.7500000	5.25000000
2	2	4	13.7500000	5.25000000

E = TYPE I SS&CP MATRIX FOR: SUBJECT(A)

DF=6	Y		Z	
Y	227.87500000	$= E(a)_{YY}$	163.00000000	$= E(a)_{YZ}$
Z	163.00000000		159.50000000	$= E(a)_{ZZ}$

Note that $\hat{\beta}_1 = E(a)_{YZ} \ / \ E(a)_{ZZ} = 163.0/159.5 = 1.022$